

# MedPanel Benchmark Runner

Compares Single MedGemma vs MedPanel across test cases.

Loading MedGemma model...

MedGemma loaded!

Loading PubMed embedding model...

Embedding model loaded!

Starting benchmark comparison...

=====

BENCHMARK: Single MedGemma vs MedPanel

=====

Running 10 test cases...

This will take approximately 10 minutes.

=====

Case 1/10

=====

Symptoms: 65 year old male. Persistent cough for 6 weeks. Night sweats, 8kg weight loss  
ov...

Ground Truth: Tuberculosis

Difficulty: HARD

♦ Testing Single MedGemma Approach...

Result: Based on the provided clinical case, the primary diagnosis is \*\*reactivation tuberculosis (TB)\*\*.

\*\*Reasoning:\*\*

\* \* ...

Time: 15.1s

CORRECT

♦ Testing MedPanel Approach...

Running Radiologist agent...

Running Internist agent...

Fetching PubMed evidence...

Running Devil's Advocate agent...

 Synthesizing final report...

 MedPanel analysis complete!

Primary Diagnosis: Tuberculosis (TB)

Panel Agreement: 95%

Escalated to Human: True

Time: 64.0s

 CORRECT

 Waiting 3 seconds before next case...

=====

Case 2/10

=====

Symptoms: 45 year old female. Severe headache for 3 days. Fever 39°C, stiff neck, photoph...

Ground Truth: Meningitis

Difficulty: HARD

=====

◆ Testing Single MedGemma Approach...

Result: Based on the clinical presentation of a 45-year-old female with severe headache for 3 days, fever (39°C), stiff neck, ph...

Time: 4.3s

 CORRECT

◆ Testing MedPanel Approach...

 Running Radiologist agent...

 Running Internist agent...

 Fetching PubMed evidence...

 Running Devil's Advocate agent...

 Synthesizing final report...

 MedPanel analysis complete!

Primary Diagnosis: Meningitis (bacterial or viral)

Panel Agreement: 90%

Escalated to Human: True

Time: 62.8s

 CORRECT

 Waiting 3 seconds before next case...

=====

Case 3/10

=====

Symptoms: 30 year old male. Sudden onset crushing chest pain radiating to left arm.

Sweati...

Ground Truth: Myocardial Infarction

Difficulty: EASY

=====

- ◆ Testing Single MedGemma Approach...

Result: Primary Diagnosis: \*\*Acute Myocardial Infarction (AMI)\*\* ...

Time: 1.5s

 CORRECT

- ◆ Testing MedPanel Approach...

 Running Radiologist agent...

 Running Internist agent...

 Fetching PubMed evidence...

 Running Devil's Advocate agent...

 Synthesizing final report...

 MedPanel analysis complete!

Primary Diagnosis: Acute Coronary Syndrome (ACS)

Panel Agreement: 90%

Escalated to Human: True

Time: 53.4s

 CORRECT

 Waiting 3 seconds before next case...

=====

#### Case 4/10

=====

Symptoms: 55 year old female. Progressive dyspnea over 3 months. Chronic dry cough.

Known ...

Ground Truth: Interstitial Lung Disease

Difficulty: MEDIUM

=====

- ◆ Testing Single MedGemma Approach...

Result: Based on the information provided, the primary diagnosis is \*\*Interstitial Lung Disease (ILD)\*\*.

\*\*Reasoning:\*\*

\* ...

Time: 15.9s

 CORRECT

- ◆ Testing MedPanel Approach...

 Running Radiologist agent...

 Running Internist agent...

 Fetching PubMed evidence...

 Running Devil's Advocate agent...

 Synthesizing final report...

 MedPanel analysis complete!

Primary Diagnosis: {'raw\_response': '\n "primary\_diagnosis": "Interstitial Lung Disease (ILD)",\n "differential\_diagnoses": [\n "Rheumatoid Lung Disease",\n "He'}

Panel Agreement: N/A%

Escalated to Human: False

Time: 73.9s

 CORRECT

 Waiting 3 seconds before next case...

---

### Case 5/10

---

Symptoms: 70 year old male. Confusion and lethargy for 2 days. Blood sugar 35 mg/dL.

Histo...

Ground Truth: Hypoglycemia

Difficulty: EASY

---

- ◆ Testing Single MedGemma Approach...

Result: Primary Diagnosis: \*\*Hypoglycemia\*\*...

Time: 1.1s

 CORRECT

- ◆ Testing MedPanel Approach...

 Running Radiologist agent...

 Running Internist agent...

 Fetching PubMed evidence...

 Running Devil's Advocate agent...

 Synthesizing final report...

 MedPanel analysis complete!

Primary Diagnosis: {'raw\_response': '\n "primary\_diagnosis": "Hypoglycemia due to insulin overdose or missed/delayed meals",\n "differential\_diagnoses": [\n "Infec'}

Panel Agreement: N/A%

Escalated to Human: False

Time: 61.9s

 CORRECT

II Waiting 3 seconds before next case...

=====

Case 6/10

=====

Symptoms: 22 year old female. Severe lower abdominal pain, right-sided. Fever, nausea, ele...

Ground Truth: Ectopic Pregnancy

Difficulty: HARD

=====

- Testing Single MedGemma Approach...

Result: Based on the provided information, the primary diagnosis is \*\*Appendicitis\*\*.

\*\*Reasoning:\*\*

\* \*\*Severe lower abdomi...

Time: 14.2s

 INCORRECT/MISSED

- Testing MedPanel Approach...

 Running Radiologist agent...

 Running Internist agent...

 Fetching PubMed evidence...

 Running Devil's Advocate agent...

 Synthesizing final report...

 MedPanel analysis complete!

Primary Diagnosis: Ruptured Ectopic Pregnancy

Panel Agreement: 95%

Escalated to Human: True

Time: 67.9s

 CORRECT

 DEVIL'S ADVOCATE MADE THE DIFFERENCE!

Single agent missed it, but MedPanel caught it!

II Waiting 3 seconds before next case...

=====

Case 7/10

=====

Symptoms: 68 year old male. Progressive weakness in legs over 6 months. Difficulty walking...

Ground Truth: Spinal Cord Compression

Difficulty: MEDIUM

=====

- Testing Single MedGemma Approach...

Result: Based on the provided information, the primary diagnosis is \*\*Spinal Stenosis\*\*.

\*\*Reasoning:\*\*

\* \*\*Progressive leg ...

Time: 17.1s

 INCORRECT/MISSED

- Testing MedPanel Approach...

 Running Radiologist agent...

 Running Internist agent...

 Fetching PubMed evidence...

 Running Devil's Advocate agent...

 Synthesizing final report...

 MedPanel analysis complete!

Primary Diagnosis: Spinal stenosis

Panel Agreement: 85%

Escalated to Human: True

Time: 58.7s

 INCORRECT/MISSED

 Waiting 3 seconds before next case...

=====

Case 8/10

=====

Symptoms: 35 year old male. High fever 40°C for 5 days. Recent travel to malaria-endemic r...

Ground Truth: Malaria

Difficulty: MEDIUM

=====

- Testing Single MedGemma Approach...

Result: Based on the provided clinical case, the primary diagnosis is \*\*Malaria\*\*.

The combination of:

\* High fever (40°C) f...

Time: 8.0s

 CORRECT

- ◆ Testing MedPanel Approach...

 Running Radiologist agent...

 Running Internist agent...

 Fetching PubMed evidence...

 Running Devil's Advocate agent...

 Synthesizing final report...

 MedPanel analysis complete!

Primary Diagnosis: {'raw\_response': '\n "primary\_diagnosis": "Malaria",\n

"differential\_diagnoses": [\n "Influenza",\n "Other infectious disease (e.g., dengue f

Panel Agreement: N/A%

Escalated to Human: False

Time: 73.3s

 CORRECT

 Waiting 3 seconds before next case...

=====

Case 9/10

=====

Symptoms: 28 year old female. Sudden severe headache, worst of life. Started 1 hour ago  
du...

Ground Truth: Subarachnoid Hemorrhage

Difficulty: HARD

=====

- ◆ Testing Single MedGemma Approach...

Result: Primary Diagnosis: \*\*Subarachnoid Hemorrhage (SAH)\*\*

Rationale:

The patient's presentation of a sudden, severe headache...

Time: 6.5s

 CORRECT

- ◆ Testing MedPanel Approach...

 Running Radiologist agent...

 Running Internist agent...

 Fetching PubMed evidence...

 Running Devil's Advocate agent...

 Synthesizing final report...

 MedPanel analysis complete!

Primary Diagnosis: {'raw\_response': '\n "primary\_diagnosis": "Cerebral venous sinus thrombosis (CVST)",\n "differential\_diagnoses": [\n "Subarachnoid hemorrhage (S

Panel Agreement: N/A%

Escalated to Human: False

Time: 66.5s

 CORRECT

 Waiting 3 seconds before next case...

=====

Case 10/10

=====

Symptoms: 50 year old male. Right upper quadrant pain for 8 hours. Pain radiates to shoulder...

Ground Truth: Cholecystitis

Difficulty: EASY

=====

- Testing Single MedGemma Approach...

Result: Based on the provided clinical case, the primary diagnosis is \*\*Biliary Colic\*\*.

\*\*Reasoning:\*\*

\* \*\*Right Upper Quad...

Time: 14.6s

 INCORRECT/MISSED

- Testing MedPanel Approach...

 Running Radiologist agent...

 Running Internist agent...

 Fetching PubMed evidence...

 Running Devil's Advocate agent...

 Synthesizing final report...

 MedPanel analysis complete!

Primary Diagnosis: Cholecystitis

Panel Agreement: 85%

Escalated to Human: True

Time: 44.0s

 CORRECT

 DEVIL'S ADVOCATE MADE THE DIFFERENCE!

Single agent missed it, but MedPanel caught it!

=====

===== FINAL RESULTS =====

 OVERALL ACCURACY:

Single MedGemma : 70.0% (7/10 correct)

MedPanel : 90.0% (9/10 correct)

Improvement : +20.0 percentage points

 DEVIL'S ADVOCATE IMPACT:

Cases where Devil's Advocate made the difference: 2

That's 20% of all cases!

===== BREAKDOWN BY DIFFICULTY =====

EASY cases (3 total):

Single : 66.7%

MedPanel : 100.0%

Improvement: +33.3 points

MEDIUM cases (3 total):

Single : 66.7%

MedPanel : 66.7%

Improvement: +0.0 points

HARD cases (4 total):

Single : 75.0%

MedPanel : 100.0%

Improvement: +25.0 points

 Results saved to: benchmark\_results.json